# Long Short Term Memory

RNN ⟶ Long Term Dependency ⟶ Vanishing Gradient Problem

① RNN ⟶ Problem

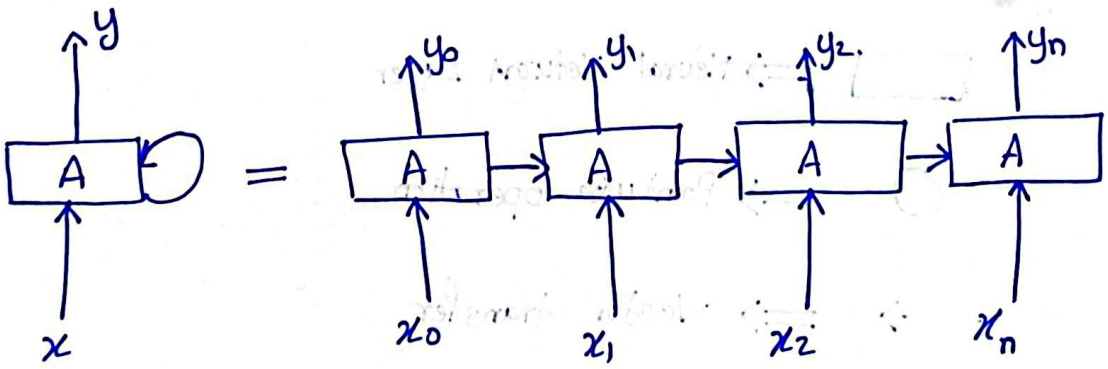
② Why LSTM

③ How LSTM
   a. Long term memory
   b. Short term memory

④ LSTM Architecture

⑤ Working of LSTM

## ⟶ Problems with RNN




Task:

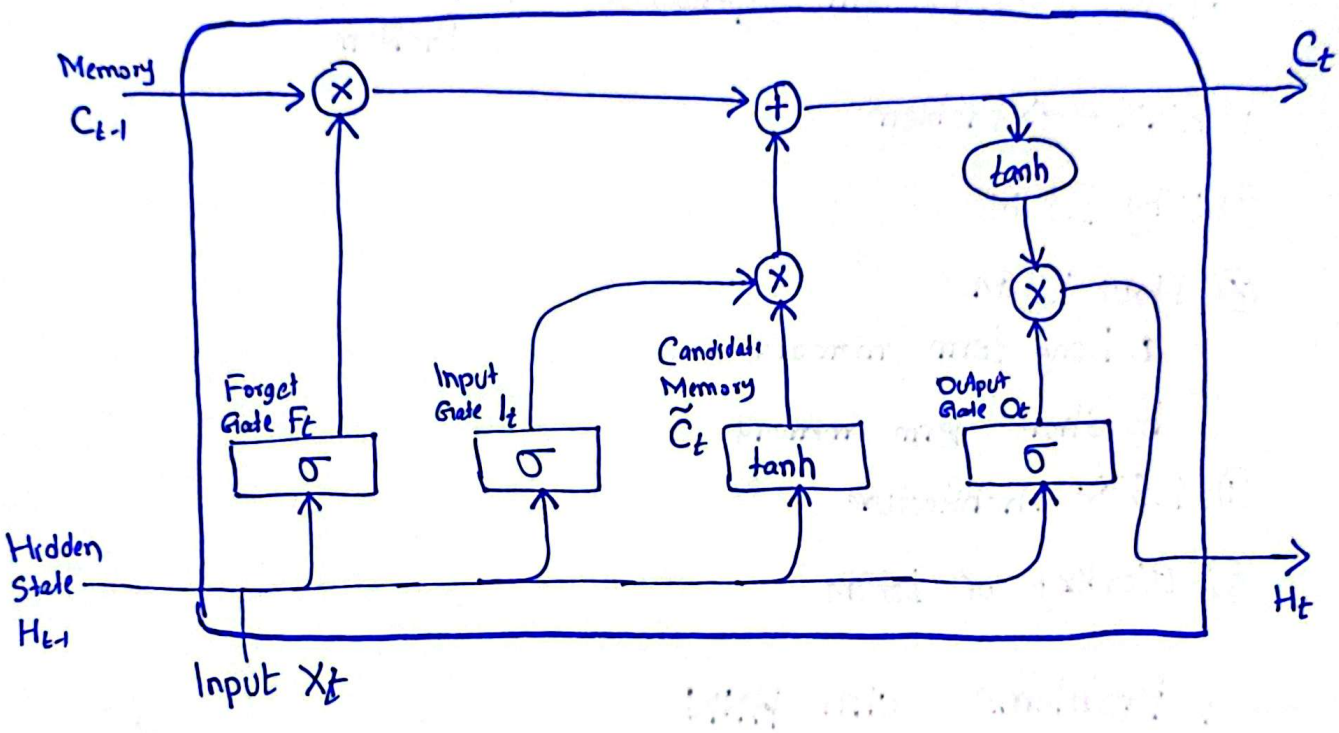
① The color of sky is blue ⟹ No Vanishing Gradient

② Much of the ML research ---------- early days of deep learning ⟹ context gap is very high

⇓

Long term depency ⟹ Vanishing Gradient

# ⟶ LSTM Architecture



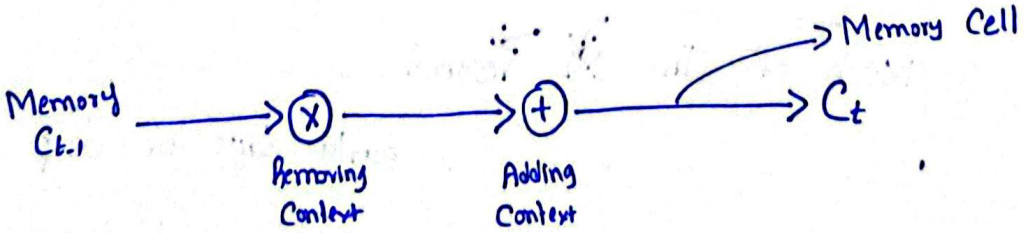

⟹ Neural Network Layer

⟹ Point wise operation

⟹ Vector transfer

⟹ Copy

⟹ Concatenate

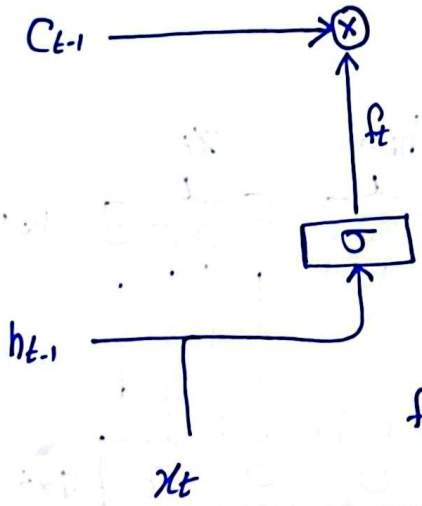


→ ~~Forget Gate~~

> Forget Gate

$h_{t-1}$ = hidden state of previous time stamp

$x_t$ = word passed as input in the current time stamp

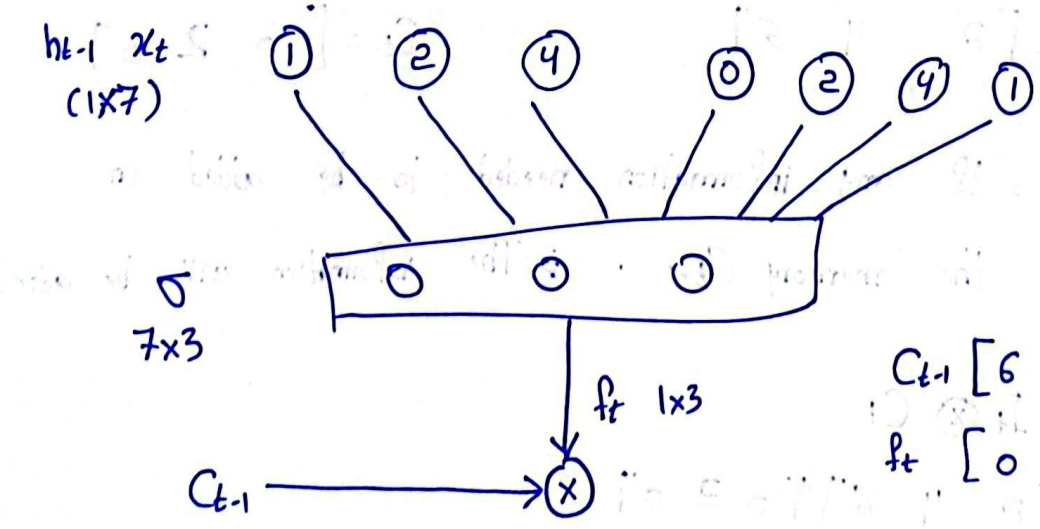



$x_t$ [0 2 4 1]

$h_{t-1}$ [1 2 4]

$C_{t-1}$ [4 2 1]

$f_t = \sigma(W[h_{t-1}, x_t] + b_f)$




$C_{t-1}$ [6 8 9]

$f_t$ [0 0 0]

[6 8 9] ⊛ [1 1 1]

= [6 8 9]

Not removing any context

[6 8 9] [0.5, 1, 0.5]

= [3 8 4.5]

Removing some context

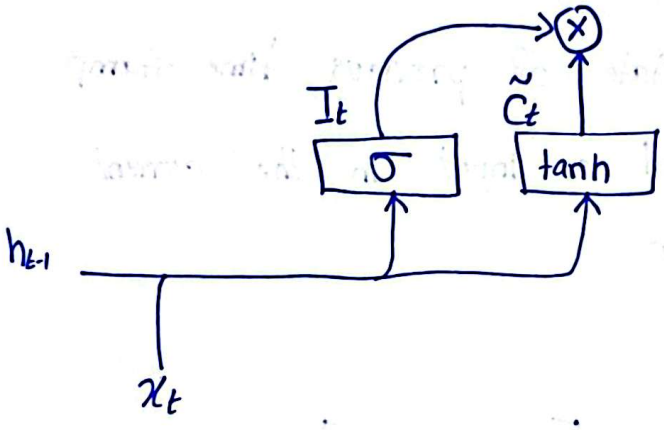
$C_{t-1} \circledast f_t$

[6 8 9] ⊛ [0 0 0]

= [0 0 0]

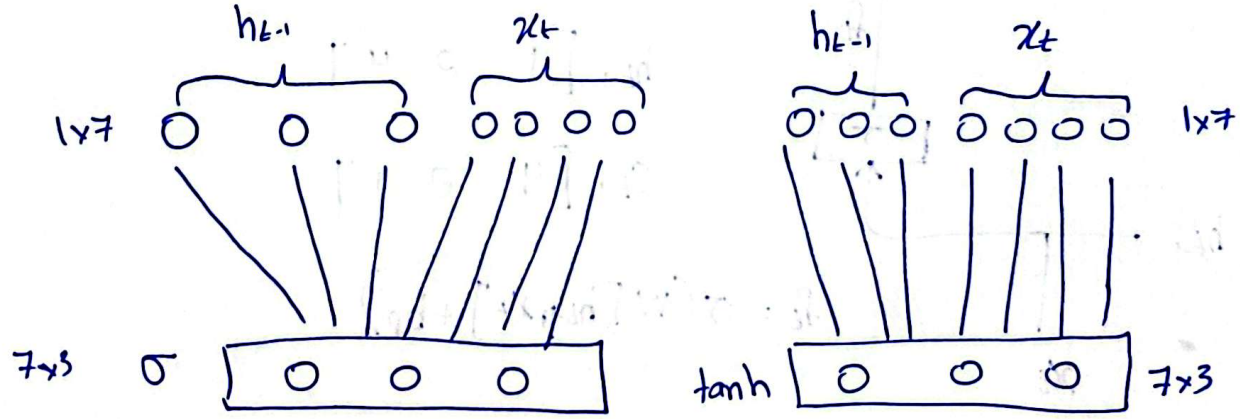
Removing all previous context

# > Input Gate and Candidate Memory




$$I_t = \sigma(W_i [h_{t-1}, x_t] + b_I)$$

$$\tilde{C}_t = \tanh(W_c [h_{t-1}, x_t] + b_c)$$




$I_t = [2 \quad 4 \quad 6]$ $\qquad$ $C_t = [0 \quad 2 \quad 0]$

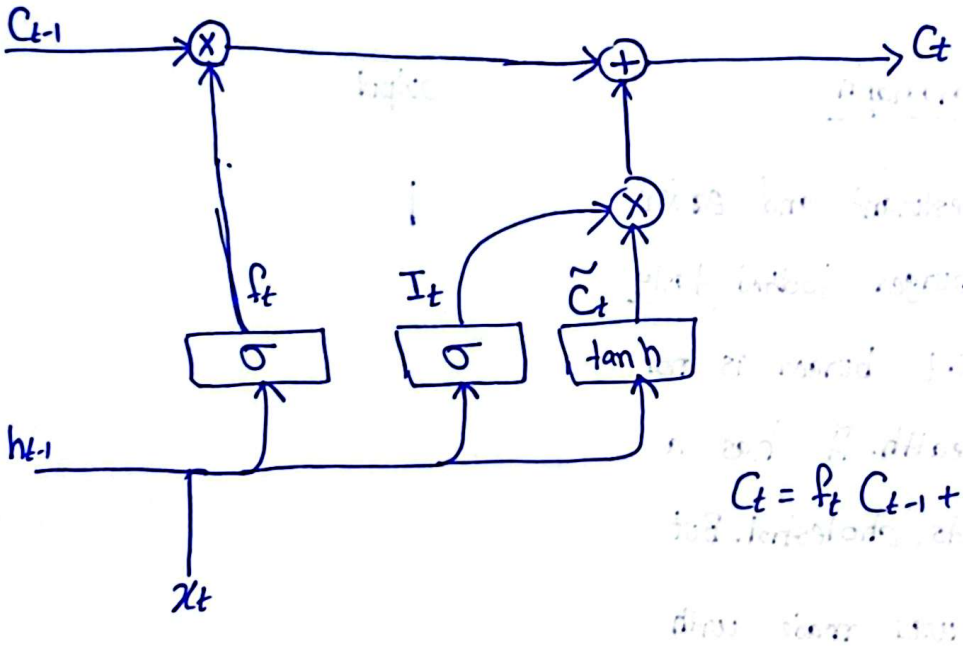
Context = If any information needed to be added in the memory $C_{t-1}$ ⟶ The information will be added

$I_t \otimes C_t$

$[2 \quad 4 \quad 6][0 \quad 2 \quad 0]$

$[0 \quad 8 \quad 0]$ ⇒ new information is added

## > Forget Gate + Input Gate and Candidate Memory

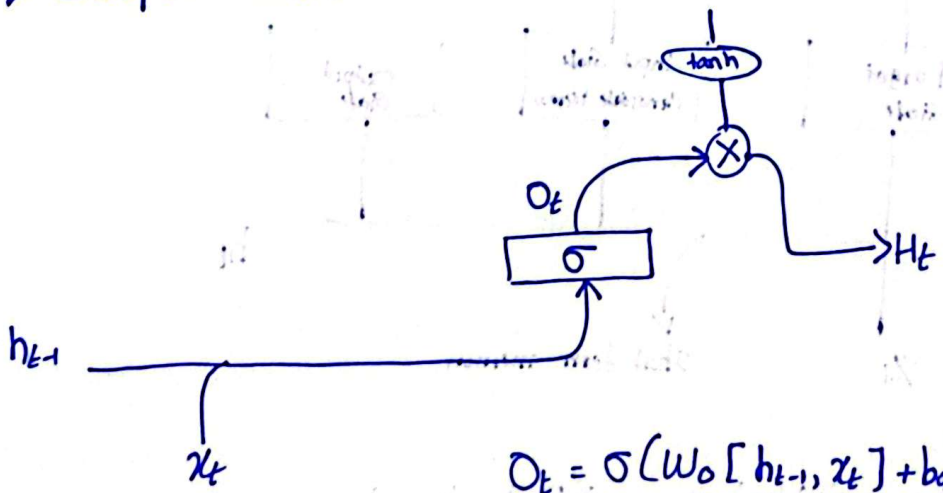



$$C_t = f_t\, C_{t-1} + I_t\, \tilde{C}_t$$

$f_t\, C_{t-1}$ $\Rightarrow$ Removing or Forgetting some memory

$I_t\, \tilde{C}_t$ $\Rightarrow$ Adding some memory through candidate memory

## > Output Gate




$$O_t = \sigma(W_o\,[h_{t-1}, x_t] + b_o)$$

$$H_t = O_t * \tanh(C_t)$$